



Europäisches Patentamt
European Patent Office
Office européen des brevets

⑪ Publication number:

0 079 739

A2

38

⑫

EUROPEAN PATENT APPLICATION

⑬ Application number: 82305926.6

⑮ Int. Cl.³: C 12 N 15/00

⑭ Date of filing: 08.11.82

C 12 N 1/00, C 12 P 21/02
C 07 H 21/04, C 07 C 103/52
//C12R1/19, C12R1/865

⑯ Priority: 12.11.81 US 320632

⑰ Applicant: THE UPJOHN COMPANY
301 Henrietta Street
Kalamazoo, Michigan 49001(US)

⑰ Date of publication of application:
25.05.83 Bulletin 83/21

⑱ Inventor: Dugaiczyk, Achilles
c/o The Upjohn Company 301 Henrietta Street
Kalamazoo Michigan 49001(US)

⑲ Designated Contracting States:
BE CH DE FR GB IT LI NL SE

⑳ Representative: Perry, Robert Edward et al,
GILL JENNINGS & EVERY 53-64 Chancery Lane
London WC2A 1HN(GB)

⑳ Albumin-based nucleotides, their replication and use, and plasmids for use therein.

⑳ The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

BEST AVAILABLE COPY

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α -fetoprotein, but the synthesis decreases drastically after birth. Recently, 10 Law et al determined the complete sequence of mouse α -fetoprotein mRNA, *Nature* 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the α -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, *Proc. Natl. Acad. Sci. USA*, 78 (1981) 3521-3525, and Gorin et al, *J. Biol. Chem.* 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the *Pst*I site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched 5 albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating 10 the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, *Msp*I(152)-*Tag*I(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T^C T C T T C T G T.....albumin mRNA
35 (3')...G A G G A A G G C G U C C m₂⁶A m₂⁶A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous 5 translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre-peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 10 384-385]. This is also a first report on the sequence of the pro-peptide for normal human serum albumin. 15

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence 20 located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the 25 human albumin mRNA (Table 1).

TABLE 1

35 30 25 20 15 10 5
 231 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu glu cys ala asp arg ala asp leu 260
 GTC TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC AAC GAA TCC TCG CTT GAA GAT TCC GAT CCT GAT GAC AGC GGC CAC CTT (890)
 261 ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys glu lys pro leu leu glu cys ala asp arg ala 265
 GGC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC ACT AAA CTC AAG GAA TCC TGT GAA AAA CCT CTC TGC CAA AAA TCR CAR TGT ATT (980)
 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 295
 GGC CAA GTC CAA AAT GAT GAG ATG CCT CCT GAC TTG CCT TCA TTA GCT GAT TTT GAA AGT AAC GAT GTC TGT AAA AAC TAT TGT ATT (1070)
 321 glu ala lys asp val phe leu gly met phe leu tyr ala arg arg His pro asp tyr ser val val leu leu ala leu ala 330
 GAG GCA AAC GAT GTC TTC GGC ATG TTT TTG GCA AGA ACC CAT CCT GAT TAC TCT GTC CTC CTC AGA CTT GCC (1160)
 351 lys thr tyr glu thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 360
 AAC ACA TAT GAA ACC ACT CTA GAG AAC TGC TGT GCT GCT GCT GCA GAT CCT CAT GAA TGC TTC GAT GAA TTT AAA CCT CCT (1250)
 381 glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu gln leu ala glu tyr lys val gln asn ala leu val arg 390
 GTC GAA GAG CCT CAG AAC TTA ATC AAA AAT TGT GAC CTT TGT CAC GAG CTC TCA AGA AAC CTA GCA AAA TGT TGT AAA CAT (1340)
 411 tyr thr lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gln val gln asn ala leu val lys 420
 TAC ACC AAC AAA GAA GTC CCA ACT CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GCA AAA TGT TGT AAA CAT (1430)
 441 pro glu ala lys arg met pro gys ala glu asp tyr leu ser val val gln leu cys val leu his glu lys thr pro val ser 450
 CCT GAA CCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC AAC CAG TTA TGT GTC CAT GAG AAA ACC CCA GTA AGT (1520)
 471 asp arg val thr lys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 480
 GAC AGA GTC ACC AAA TCC TGC ACA GAA TCC TTG GTG AAC CAG CCA CCA TAC GTC GAT GAA ACA TAC GTT CCC AAA (1610)
 501 glu phe asn ala glu thr phe thr phe his ala asp lle cys thr leu ser gln lys arg aln lle lys lys aln thr ala leu val 510
 GAC TGT AAA CCT GCA GAT ATA TGC ACA CTT TCT GAG AAC GAG AGA CAA CAA ATC AAG AAA ACT GCA CTT CCT GTT (1700)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and 10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 15 680-685.

Example 2 Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Bolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [³²P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pH A36 and pH A206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris-HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml 5 chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of 10 L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and 20 Boyer, H.W. (1974) *J. Virol.* 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) *Biochemistry* 11, 1242-1250] gels.

Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline 25 phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ [³²P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and 30 Gilbert, W. (1980) *Methods Enzym.* 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) *FEBS Letters* 87, 107-110].

Example 5 Recombinant Plasmids pH A36 and pH A206

35 As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pH A36 contained the largest insert of an albumin cDNA sequence. Both plasmids pH A36 and pH A206 have been deposited in a viable *E. coli* host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5 HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. ~~upon the grant of a patent. It should be understood that the availability~~
10 of these deposits does not constitute a license to practice the subject invention in derogation of ~~patent rights granted with the subject instrument by governmental action.~~

15 E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.
20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

25 Assembling the pieces together is a straightforward task of restriction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the
30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

35 (a) Escherichia coli

(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the 5 existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one 10 of the yeast plasmid vectors, e.g., YEp6, at the Eco R1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.R. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed *supra*.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning 20 of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

Example 8 Screening of Clones Producing Albumin

35 Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies 5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. 10 Peters: *Purification and Properties of Serum Albumin*, in: *The Plasma Proteins*, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pH A36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pH A206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

0079739

4083

-13-

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230

-1 -6 p r o -1 1
ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his lys asp leu arg phe lys asp leu arg phe lys glu asp leu arg phe lys
TCC GCT TAT TCC AGC CGT GTC ATT GGC TTT GCT CAC CAT GCA CAC GAG CTC GAT CGT CAT CGG TTT AAA GAT TTC GCA GAA AAT TTC AAA (170)

21 ala leu val ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
GGC TTG CGC TTC ATT GGC TTT GCT CAC TAT CTC CAG CAC TAT GCA GAT CAT GTC AAA TTA GCA ACT CAA TTT GCA (260)

51 53 60 62 70 75 80
lys thr oys val ala esp glu ser ala glu asn cys esp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTC CCT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CCA GAC AAA TTA TCC ACA GTC GCA ACT CTT (350)

81 91
arg glu thr tyr gly glu met ala esp cys cys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys asp asp asn pro
CGT GAA ACC TAT GGT GAA ATG GCT GAC TGC TGT GCA AAA CAA GAA CCT GCG AGA AAT GAA TGC TTC TGC CAA CAC AAA GAT GAC AAC CCA (440)

111 120 124 130 140
asn leu pro arg leu val phe tyr ala pro glu val met cys thr ala phe his esp asn glu glu phe leu lys tyr leu try
AAC CTC CCC CGA TTG CGC AGA CCA GAG GTC ATT GAT GTC ATT GCT TTT CAT GAC AAT GAA GCG ACA TTT TGC AAA AAA TAC TTA TAT (530)

141 150
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu phe phe ala lys arg tyr lys ala ala phe thr alu cys cys aln
CAA ATT GCC AGA AGA CAT CCT TAC TAT GCC CCC GAA CTC CTT CCT TTT GCT TTT GCT TAT AAA AGC TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)

171 177 180 190 200
ala ala esp lys ala ala cys leu leu pro lys leu esp glu leu arg esp glu gly lys ala ser ser ala lys qln arg leu lys cys
GCT GAT AAA CCT GCC TGC CTC AGC CTC GAT GAA CTT CGG GAT GAA CCT GCG AAC GCT TCC TCT CCC AAA CAC AGA CTC AAC TGC (710)

201 210 220 230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg ala phe ala glu
GCC AGT CTC CAA AAA TTT CGA GAA ACA CCT TTC AAA GCA TGC GCA GAG CTC AGC AAC GCA TTT CCC AAA CCT GAG TTT GCA GAA (300)

-14-

0079739
4083

-15-

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

1	10	20
21	30	40
51	60	70
81	90	100
111	120	130
141	150	160
171	177	180
201		
25	15	5
30	20	10
35		

asp ala his lys ser glu val ala his arg phe lys asp leu ala glu glu asn phe lys
 GAT GCA GAC AAG ACT GAG CTT CAT CGG TTT AAA GAT TTC GCA GAA AAT TTC AAA (170)

 ala leu val leu lle ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
 GCC TTG CTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GAA AAA TTA GTG AAT GAA GTT ACT CAA TTT GCA (260)

 lys thr cys val ala asp glu ser ala glu asn oys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
 AAA ACA ACC TAT CCT GCT GAT GAG TCA CCT GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CCT TTT CGA GAC AAA TTA TGC ACA ACT CTT (350)

 arg glu thr tyr gly glu met ala asp cys oys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys asp asp asn bro
 CCT GAA ACC TAT CCT GAA ATG CCT GAC TCC TGT GCA ATT GCA TAT GAA TCC TTC TGT CAA CAC AAA GAT GAC AAC CCA (440)

 asn leu pro arg leu val arg pro glu val met cys thr ala phe his asp asn glu glu thr phe leu lys tyr leu try
 AAC CTC CCC CCA TTG CTG AGA CCA ATT GCA TAT GAT GCG ATG TGC ACT CCT TTT CAT GAC AAT GAA TTT TTG AAA AAA TAC TTA TAT (530)

 glu lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala ala phe thr glu cys cys aln
 GAA ATT GCC AGA AGA CAT CCT TAC TAT GGC CCC GAA CTC CTT TTT GCT ATT AAA AGC TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)

 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys aln arg leu lys oys
 CCT GCT GAT AAA GCT GCC TCC CTC CCA AAG CTC GAT GAA CTT CGG CAT GAA CGG AAG CCT TCG TCT GCC AAA CAG AGA CTC AAC TGT (710)

 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu
 CCC AGT CTC CAA AAA TTT CGA GAA AGA CCT TCC AAA CGG CTC AGC CAG AGA TTT CCC AAA CCT GAC TTT GCA GAA (800)

35 30 25 20 15 10 5

231 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu glu cys ala asp arg ala asp leu
GTT TCC AAC TTA GTC ACA ACC CTT ACC GAA GTC CAC ACG GAA TCC GAT CTC CCT GAT GAC GAT CTC CTT GAA TGT CCT GAT GAC GTC CTT (880)

261 265 ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys cys glu lys pro leu glu lys ser his cys lle
CCC GAA GTC GAA AAT CAA GAT TCC AGT AAA CTC AAG GAA TCC TGT GAT CCT GAT GTC CCT GAA AAA TCC TAC GAC TCC ATT (980)

291 300 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala
CCC GAA GTC GAA AAT GAT GAG ATC ACC CCT CCT GAC TTG CCT TCA TTA GCT GAT TTT GTC TGT GAA AGT AAC CAT GTT TCC AAA AAC TAT CCT (1070)

321 330 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg his pro asp tyr ser val val phe asp glu ser lys pro leu ala
GAG GCA AAC GAT GTC TTC TTG GGC ATC TTT TTG TAT GAA AGA ACC CAT CCT GAT TAC TCT GTC CTC CTC GCA CCT GTC CCC (1160)

351 360 361 lys thr tyr glu thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
AGC ACA TAT GAA ACC ACT CTA GAG AAC TGC TGT CCC CCT GCA GAT CCT CCT GAA TCA TCC TAT GCA TAC TCT GCA TAA TTC CAG AAT GGC CTC TTG TGT CCT (1250)

381 390 392 400 410 val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu aln leu qly glu tyr lys phe gln asn ala leu leu val arg
GTC GAA GAC CCT CAG CAC TCA ACT AAA CAA AAT TTA ATC AAC TCT GAC CTT TTT GAG CAG CCT GCA CAG TAC TCA AAA TTC CAG AAT GGC CTC TTG TGT CCT (1340)

411 420 430 440 450 460 461 470 480 490 500 510 520 530

tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val gln lys thr pro val ser
TAC ACC AAC AAA GCA CCC CAA GTC TCA ACT CCA ACT CTT GTA GAC GTC TCA AGC AAC CTA GCA AAA GTC GGC ACC AAA TGT TGT AAA CAT (1430)

441 448 450 460 461 470 480 490 500 510 520 530

pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu asn gln leu cys val his glu val asp ala thr tyr val pro lys
CCT GAA GCA GAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG CGA CCA TGC TGT TCA GCT GTC GAA ACA TAC GCA CCT CCC GTC ACT (1520)

471 476 477 480 490 500 510 520 530

asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val his lys aln thr ala leu val
GAC AGA GTC ACC AAA TGC TCC ACA GAA TCC TTG GTC AAC AGG CGA CCA CAA ATC AAC GAG AGA CAA ACA TAC GCA CCT GTC GAA (1610)

501 510 520 530

glu phe asn ala glu thr phe thr phe his ala asp lle cys thr leu ser qly lys glu arg aln lle lys lys aln thr ala leu val
GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAC TCC TGT GTC AAC CAA ACT GCA CCT GTC GAA (1700)

5

10

15

20

25

30

35

40

45

50

55

60

65

70

75

80

85

90

95

100

105

110

115

120

125

130

135

140

145

150

155

160

165

170

175

180

185

190

195

200

205

210

215

220

225

230

235

240

245

250

255

260

265

270

275

280

285

290

295

300

305

310

315

320

325

330

335

340

345

350

355

360

365

370

375

380

385

390

395

400

405

410

415

420

425

430

435

440

445

450

455

460

465

470

475

480

485

490

495

500

505

510

515

520

525

530

535

540

545

550

555

560

565

570

575

580

585

590

595

600

605

610

615

620

625

630

635

640

645

650

655

660

665

670

675

680

685

690

695

700

705

710

715

720

725

730

735

740

745

750

755

760

765

770

775

780

785

790

795

800

805

810

815

820

825

830

835

840

845

850

855

860

865

870

875

880

885

890

895

900

905

910

915

920

925

930

935

940

945

950

955

960

965

970

975

980

985

990

995

1000

1005

1010

1015

1020

1025

1030

1035

1040

1045

1050

1055

1060

1065

1070

1075

1080

1085

1090

1095

1100

1105

1110

1115

1120

1125

1130

1135

1140

1145

1150

1155

1160

1165

1170

1175

1180

1185

1190

1195

1200

1205

1210

1215

1220

1225

1230

1235

1240

1245

1250

1255

1260

1265

1270

1275

1280

1285

1290

1295

1300

1305

1310

1315

1320

1325

1330

1335

1340

1345

1350

1355

1360

1365

1370

1375

1380

1385

1390

1395

1400

1405

1410

1415

1420

1425

1430

1435

1440

1445

1450

1455

1460

1465

1470

1475

1480

1485

1490

1495

1500

1505

1510

1515

1520

1525

1530

1535

1540

1545

1550

1555

1560

1565

1570

1575

1580

1585

1590

1595

1600

1605

1610

1615

1620

1625

1630

1635

1640

1645

1650

1655

1660

1665

1670

1675

1680

1685

1690

1695

1700

1705

1710

1715

1720

1725

1730

1735

1740

1745

1750

1755

1760

1765

1770

1775

1780

1785

1790

1795

1800

1805

1810

1815

1820

1825

1830

1835

1840

1845

1850

1855

1860

1865

1870

1875

1880

1885

1890

1895

1900

1905

1910

1915

1920

1925

1930

1935

1940

1945

1950

1955

1960

1965

1970

1975

1980

1985

1990

1995

2000

2005

2010

2015

2020

2025

2030

2035

2040

2045

2050

2055

2060

2065

2070

2075

2080

2085

2090

2095

2100

2105

2110

2115

2120

2125

2130

2135

2140

2145

2150

2155

2160

2165

2170

2175

2180

2185

2190

2195

2200

2205

2210

2215

2220

2225

2230

2235

2240

2245

2250

2255

2260

2265

2270

2275

2280

2285

2290

2295

2300

2305

2310

2315

2320

2325

2330

2335

2340

2345

2350

2355

2360

2365

2370

2375

2380

2385

2390

2395

2400

2405

2410

2415

2420

2425

2430

2435

2440

2445

2450

2455

2460

2465

2470

2475

2480

2485

2490

2495

2500

2505

2510

2515

2520

2525

2530

2535

2540

2545

2550

2555

2560

2565

2570

2575

2580

2585

2590

2595

2600

2605

2610

2615

2620

2625

2630

2635

2640

2645

2650

2655

2660

2665

2670

2675

2680

2685

2690

2695

2700

2705

2710

2715

2720

2725

2730

2735

2740

2745

2750

2755

2760

2765

2770

2775

2780

2785

2790

2795

2800

2805

2810

2815

2820

2825

2830

2835

2840

2845

2850

2855

2860

2865

2870

2875

2880

2885

2890

2895

2900

2905

2910

2915

2920

2925

2930

2935

2940

2945

2950

2955

2960

2965

2970

2975

2980

2985

2990

2995

3000

3005

3010

3015

3020

3025

3030

3035

3040

3045

3050

3055

3060

3065

3070

3075

3080

3085

3090

3095

3100

3105

3110

3115

3120

3125

3130

3135

3140

3145

3150

3155

3160

3165

3170

3175

3180

3185

3190

3195

3200

3205

3210

3215

3220

3225

3230

3235

3240

3245

3250

3255

3260

3265

3270

3275

3280

3285

3290

3295

3300

3305

3310

3315

3320

3325

3330

3335

3340

3345

3350

3355

3360

3365

3370

3375

3380

3385

3390

3395

3400

3405

3410

3415

3420

3425

3430

3435

3440

3445

3450

3455

3460

3465

3470

3475

3480

3485

3490

3495

3500

3505

3510

3515

3520

3525

3530

3535

3540

3545

3550

3555

3560

3565

3570

3575

3580

3585

3590

3595

3600

3605

3610

3615

3620

3625

3630

3635

3640

3645

3650

3655

3660

3665

3670

3675

3680

3685

3690

3695

3700

3705

3710

3715

3720

3725

3730

3735

3740

3745

3750

3755

3760

3765

3770

3775

3780

3785

3790

3795

3800

3805

3810

3815

3820

3825

3830

3835

3840

3845

3850

3855

3860

3865

3870

3875

3880

3885

3890

3895

3900

3905

3910

3915

3920

3925

3930

3935

3940

3945

3950

3955

3960

3965

3970

3975

3980

3985

3990

3995

4000

4005

4010

4015

4020

4025

4030

4035

4040

4045

4050

4055

4060

4065

4070

4075

4080

4085

4090

4095

4100

4105

4110

4115

4120

4125

4130

4135

4140

4145

4150

4155

4160

4165

4170

4175

4180

4185

4190

<p style="text-align: right

0079739

4083

4083
7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

ser	ala	tyr	ser	erg	gly	val	phe	arg	arg
CCG	CTT	TTG	TAI	TCC	ACG	GGT	GTC	TTT	CCA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

35	30	25	20	15	10	5		
-6	p r o	-1						
arg gly val phe arg arg asp ala his lys ser glu val ala his	lys arg phe lys asp leu ala glu glu asp	lys	20					
AGG GGT GTC TTT CGT CCA GAT CGA CAC AAG AGT GAC GTC GCT GAT CGG TTT AAA GAT TTT GCA GAA AAT TTC AAA	(170)							
21	ala leu val leu ile ala phe ala gln tyr leu gln gln oys pro phe glu asp his val lys leu val asn glu val thr glu phe ala	50						
GCC TTG CTC ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GTC AAA TTA GTC AAT GAA GAT CAA TTT GCA	(260)							
51	53	55	60	62	64	70	75	80
lys thr oys val ala asp glu ser ala glu asn oys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu	100	101	102	103	104	105	106	107
AAA ACA TGT GCT GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA CCT GCA ACC TCA ACC CCT GCT GTC AGA AAT GAA TCA GTC TAC GAA CAC AAA GAT GAC AAC CCA	(440)							
81	90	91	92	93	94	95	96	97
arg glu thr tyr gly glu met ala asp oys cys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys asp asp asn pro	100	101	102	103	104	105	106	107
CCT GAA ACC TAT CCT GGT GAA ATG CCT GAC TGC TGT GCA AAA CAA CCT GCA GAC TAA GAT GAA TCA TTT TGC TAC TTA TAT	(330)							
111	120	124	125	126	127	128	129	130
asn leu pro arg leu val arg pro glu val met cys thr ala phe his asp asn glu glu thr phe leu lys lys tyr leu try	140							
AAC CTC CCC CGA TTG GTC AGA CCA GAG GTT GAT GTC ATG TGC ACT CCT GAC TAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT	(330)							
141	150	154	155	156	157	158	159	160
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys ala lys arg tyr lys ala ala phe thr glu oys cys qin	168	169	170	171	172	173	174	175
CAA ATT GCC AGA CAT CCT TAC TTT TAT GGC CCC GAA CTC CTT TTC ATT CCT AAA AGC TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA	(620)							
171	177	180	183	186	189	192	195	200
ala ala asp lys ala ala oys leu leu pro lys leu asp glu lys ala ser ser ala lys ala ser ser ala lys ala ser ser ala lys cys	200							
GCT CCT GAT AAA GCT GCT GTC CTC CTC GAT GAA CTC AAG CTC TGC TGC CTC GAT GAA CGG CAT GAA CGG AAC GCT TCG TCT GCA GTC AAC TCT	(710)							
201	210	214	218	222	226	230	234	238
ala ser leu gln lys ala phe gly glu arg ala phe lys ala val ala arg leu ser gln arg ala phe lys ala glu phe ala glu	238							
GCC ACT CTC CAA AAA TTT GCA GAA AGA CCT TGC CCC AAA CCT GCA GTC AGC CAG AGA TTT CCC AAA CCT GAG TTT GCA GAA	(330)							

-21-

0079739
4083

5

10

15

20

25

30

35

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	10	15	20	25	30	35	
-1	-6	-1	-1	-1	-1	-1	-1
ser	ala	tyr	ser	arg	gly	val	phe
TCC	GCT	TAT	TCC	AGC	GGT	TTC	TTC

-16 p r o -10

Met lys trp val thr phe ile ser leu leu phe leu ale ser

ATG AAG TGG GTC ACC TTT ATT TCC CTT CTT CTC TTT AGC (30)

21 30 34 40 45 50

ala leu val ile ale phe ale gln tyr leu gln gln oys pro phe glu asp his val lys leu val asn glu val thr glu phe ale

GCC TTC GTC TTC ATT GGC TTT CCT CAG TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GAT CAT GTC AAA TTA GTG AAT GAA GTC ACT GAA TTT GCA (126n)

51 53 60 62 70 75 80

lys thr cys val ale asp glu ser ale glu asn cys esp lys ser leu his thr leu phe gly asp lys leu cys thr val ale thr leu

AAA ACA TGT GTC ATT GGT GAT GCA AAT TGT CAC TCA GCT CAT ACC CTT CTC AAA GAC AAA TTA TCC ACA GTC GCA ACT CTT (35n)

81 90 91 100 101 110

arg glu thr tyr gly glu met ale esp cys cys ale lys gln glu pro gly arg asn glu oys ale leu gln his lys asp asp asn pro

CGT GAA ACC TAT GGT GAA ATG GCT GAC TGC TGT GCA AAA CAA GAA CCT CCC AGA AAT GAA TCC TTC TGC CAA CAC AAA CAT GAC AAC CCA (46n)

111 120 124 130 140

asn leu pro arg leu val ale esp val met cys thr ale phe his asp asn glu glu thr phe leu lys tyr leu try

AAC CTC CCC CCA TTC GTC AGA CCA GAG GTT GAT GTC ATG TCC ACT CCT TTT CAT GAC AAT GAA GAG ACA TTT TGC AAA AAC TAC TTA TAT (33n)

141 150 160 168 169 170

glu ile ale arg his pro tyr phe tyr ale pro glu leu leu phe ale lys arg tyr lys ale phe thr glu cys cys qln

CAA ATT GGC AGA AGA CAT CCT TAC TTT TAT GGC CGG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT TTT AGA TGA TGT TGC CAA (62n)

171 177 180 190 200

ala ale esp lys ale ale oys leu leu pro lys leu esp glu glu arg esp glu qly lys ale ser ser ale lys ale leu lys cys

CCT CCT GAT AAA CCT GCC TCC CTC TGC TGT CCA AAC CTC GAT GAA CTT CGG GAT GAA GCT GAA CCC AAC CCT TCG TCT GCT GCA CTC AAC TGC (71n)

201 210 220 230

ala ser leu gln lys phe gly glu arg ale phe lys ale ale trp ale val ale arg leu ser gln arg phe pro lys ale glu phe ale glu

CCC AGT CTC CAA AAA TTT GCA GAA AGA CCT TCC AAA GCA TGC AGA CCT CCC CTC ACC CAG ACA TTT CCC AAA GCT GAC TTT GCA GAA (33n)

0079739
4083

-25-

5

10

15

20

25

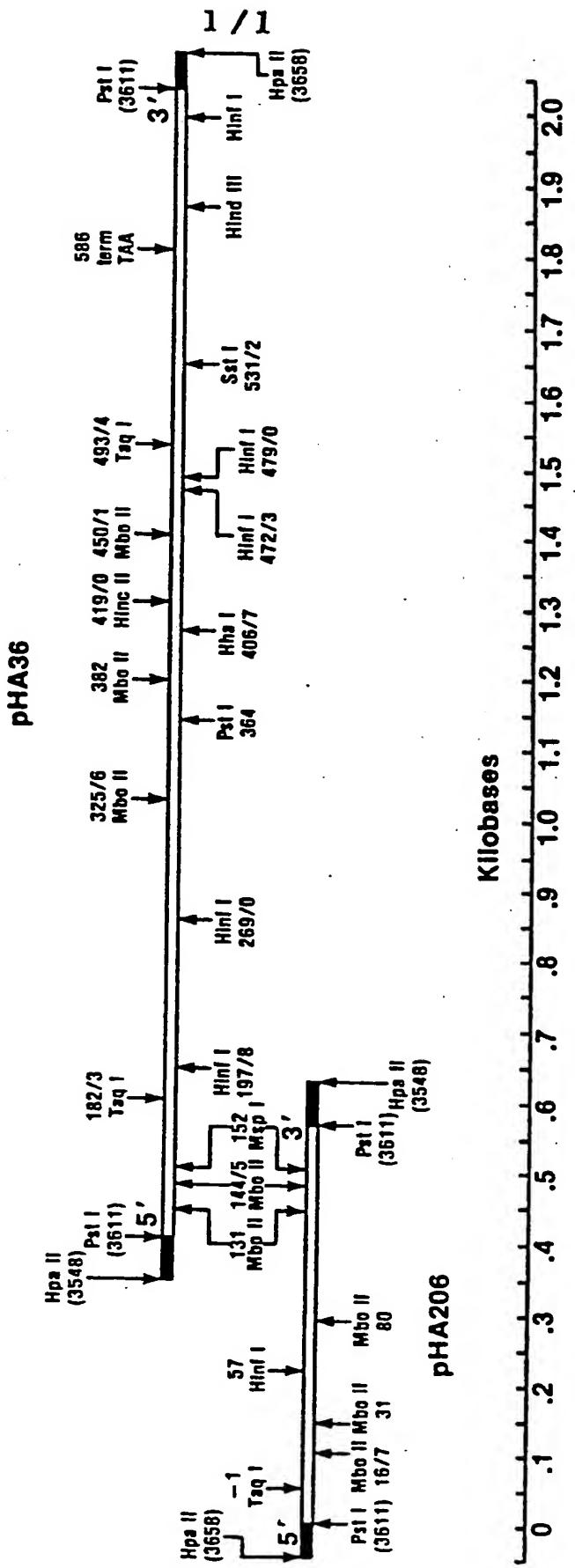
30

35

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

0079739

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.